Contemporary Clinical Trials Communications 3 (2016) 65–69

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Robust inference for responder analysis: Innovative clinical trial design using a minimum p-value approach

Yunzhi Lin

Takeda Development Center Americas, Inc., Deerfield, IL 60015, USA

ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form
31 March 2016
Accepted 4 April 2016
Available online 14 April 2016

Keywords:
Clinical trials
Responder analysis
Multiple testing
Robust inference

ABSTRACT

Responder analysis is in common use in clinical trials, and has been described and endorsed in regulatory guidance documents, especially in trials where "soft" clinical endpoints such as rating scales are used. The procedure is useful, because responder rates can be understood more intuitively than a difference in means of rating scales. However, two major issues arise: 1) such dichotomized outcomes are inefficient in terms of using the information available and can seriously reduce the power of the study; and 2) the results of clinical trials depend considerably on the response cutoff chosen, yet in many disease areas there is no consensus as to what is the most appropriate cutoff. This article addresses these two issues, offering a novel approach for responder analysis that could both improve the power of responder analysis and explore different responder cutoffs if an agreed-upon common cutoff is not present. Specifically, we propose a statistically rigorous clinical trial design that pre-specifies multiple tests of responder rates between treatment groups based on a range of pre-specified responder cutoffs, and uses the minimum of the p-values for formal inference. The critical value for hypothesis testing comes from permutation distributions. Simulation studies are carried out to examine the finite sample performance of the proposed method. We demonstrate that the new method substantially improves the power of responder analysis, and in certain cases, yields power that is approaching the analysis using the original continuous (or ordinal) measure.

1. Introduction

In many disease areas in which "hard" clinical endpoints such as mortality are not appropriate measures of efficacy, rating scales and other continuous measures are used for the evaluation of treatments. For instance, in schizophrenia clinical trials, the MATRICS Consensus Cognitive Battery (MCCB) or the Negative Symptom Assessment-16 (NSA-16) are frequently used instruments for evaluating psychopathology in study subjects. Other examples include the use of the Expanded Disability Status Scale (EDSS) in multiple sclerosis trials, the use of exercise tolerance (ET) measures in trials of heart failure therapies, and etc. In such studies, overall treatment effect has typically been tested by assessing the difference in mean change over time of the continuous (or ordinal) measure between the treatment and control group. Although such analyses are usually the primary outcomes, one problem is that the translation of the results into clinical practice is difficult. We might not know what, for example, a difference which is statistically significant but amounts to only 1 MCCB point in magnitude means from a clinical perspective. Such a problem can be addressed by using a responder analysis, in which each subject is classified as either a "responder" or a "non-responder", and the proportions of patients who benefit are quantified and compared between treatment groups. A common approach is to define a threshold for the change from baseline in the continuous (or ordinal) endpoint, and define a patient as a "responder" if his/her change value is above (or below) the threshold.

Responder analysis provides several benefits and hence is in many cases proposed or recommended by regulatory guidance or clinical communities to be used in clinical trials. For example, draft guidance from the FDA on patient-reported outcomes specifically endorsed the responder analysis as an alternative approach to assessing clinical relevance [1]. The procedure is useful, because responder rates can be understood more intuitively than a difference in means of rating scales. It also helps ensure that a reported statistically significant result represents a clinically meaningful benefit. However, two major issues arise from this procedure. First, it is well known that dichotomization tends to result in a loss of statistical power compared to an analysis of the original continuous variable. The procedure hence is inefficient in terms of using the information available and requires greater sample size in clinical

E-mail address: yzlinn@gmail.com.

http://dx.doi.org/10.1016/j.conctc.2016.04.001

trials as analyzed, for example, by Snapinn and Jiang [2]. The second issue with responder analysis is that the results of clinical trials depend considerably on the response cutoff chosen. Yet in many disease areas across different clinical trials, various definitions of response have been used, and there is no consensus as to which is the most appropriate one [3]. If a cutoff is chosen post hoc, this is potentially an inappropriate manipulation of the data.

The issues and challenges inherent in the responder approach deserve particular attention in the development and licensing of new therapeutics. The present paper addresses the two issues mentioned above, offering a novel approach for responder analysis that could both improve the efficiency and power of responder analysis and explore different responder cutoffs if an agreed-upon common cutoff is not present.

Pre-specification of the responder cutoff and a properly planned statistical analysis are essential to avoid multiple comparisons and inflated type I error rates. But how can we pre-specify when we are not certain which responder cutoff is the optimal one? Ganju et al. recently proposed to analyze clinical trial data by pre-specifying multiple test statistics and using a combined statistic – the minimum p-value – for inference when there is uncertainty about what candidate primary endpoint, hypothesis, or statistical test to use in planning a clinical trial [4–7]. The critical value for hypothesis testing comes from permutation which consists of re-randomizing the treatment assignments and calculating the combined statistic. For instance, for a trial with a time-to-event endpoint, it might be unclear at the planning stage of the trial whether a log-rank test or a stratified log-rank test would be more appropriate for the analysis. Using the proposed method, the trialists can pre-specify both tests and use the minimum of the p-values as the new test statistic. It has been shown that the method, referred to as MinP, is robust, controls the type I error rate, and provide statistical power that is closest to the best-performing statistic.

In this paper, we borrow the idea from Ganju et al. and extend the use of MinP to clinical trials analyzed by the responder approach. We propose a statistically rigorous clinical trial design that pre-specifies multiple tests of responder rates between treatment groups based on a range of pre-specified responder cutoffs, and uses the minimum of the p-values for formal inference. The null hypothesis associated with the multiple tests is that there is no treatment effect however the "responder" is defined. The alternative hypothesis is that there is a significantly greater proportion of responders in the new treatment group, with the criterion for "responding" being one of the pre-specified cutoffs. The proposed method therefore provides not only a formal test for the treatment effect, but also an estimate of the optimal responder cutoff, which could be carried forward into future trials. More importantly, we show that the proposed method, which we will refer to as MinP responder analysis in the rest of this paper, substantially improves the power of responder analysis. In many cases, the MinP responder analysis yields power that is approaching the analysis using the original continuous (or ordinal) measure.

The rest of the paper is structured as follows. In Section 2, we describe the proposed method. The method is then illustrated on a real data example in Section 3, and simulation studies evaluating the performance of the MinP responder analysis are presented in Section 4. Discussions and conclusion are given in Section 5.

2. Method

2.1. Design considerations

In general, suppose that the clinical endpoint is a continuous variable, Y, such that larger values represent better efficacy. Note that Y could represent a measurement taken at the conclusion of the trial or a change in that measurement from its baseline value. Assume, without loss of generality, a two-treatment trial with N_A subjects randomized to treatment A (e.g. experimental treatment) and N_B to treatment B (e.g. control). There is interest in the mean difference in this endpoint, μ, between the experimental treatment and the control.

The difference in treatment effects can be determined using the original continuous scale. In this case, the typical null hypothesis (assuming one-sided testing) is that of no difference, or $\mu \leq 0$, versus the alternative hypothesis $\mu > 0$.

Alternatively, with responder analysis, a threshold value is defined above which a subject is considered to be a "responder", and below which a subject is considered to be a "non-responder". If we let y_0 represent the threshold value, then

$$W = \begin{cases} 1 & \text{if } Y \geq y_0 \\ 0 & \text{if } Y < y_0 \end{cases}$$

is a binary variable indicating whether or not the subject is a responder. Now let p_A and p_B be the response rates in the experimental group and the control group, respectively. Therefore the null hypothesis for the responder analysis is $p_A \leq p_B$, and the alternative is $p_A > p_B$. If the responder null hypothesis is rejected then both statistical significance and clinical relevance are concluded. When the responder cutoff value y_0 is not well established and properly validated before the study, however, the results from such responder analysis could be inadequate or irrelevant. Moreover, as pointed out before, this approach substantially reduce the power of the study as information is lost through dichotomization the continuous endpoint.

2.2. MinP responder analysis procedure

Consider a setting for which there is a lack of consensus on the proper responder cutoff to use. Without loss of generality, assume that the continuous endpoint (and hence the responder cutoff y_0) take values in the interval (0, 100). The objective of the proposed design is to

1. Formally test for any treatment effect, i.e. determining whether a significantly greater proportion of subjects in the experimental arm "respond" to the treatment compared to the control arm based on a certain responder cutoff; and
2. Identify optimal responder cutoff which could be carried forward into future trials.

As the responder cutoff point is not well-established, we design the trial by pre-specifying multiple tests of responder rates between treatment groups based on a range of pre-specified responder cutoffs. Based on prior medical knowledge and discussion with the clinical team, a series of plausible candidate responder cutoffs $\{y_{0,k}: k = 1, 2, \ldots, K\}$ in the interval (0, 100) can be pre-specified. For instance, $\{y_{0,k}\} = \{10, 20, 30, \ldots, 90\}$. For each candidate cutoff $y_{0,k}$, a proportion test T_k will be performed to test the null hypothesis that $p_A \leq p_B$, resulting in a series of p-values $\{p_k\}$. A natural approach to converting a series of p-values that are calculated over the range of possible cutoff values into a single statistic is then to take the minimum:

$$minP = \min(p_1, \ldots, p_K)$$

Because of the well-known multiple testing problem, the standard asymptotic theory does not apply to the new statistic, *minP*. To provide a statistically valid p-value for *minP*, we propose to use the permutation distribution of *minP*, in which the treatment group